Big Data and Computing Visions

www.bidacv.com

Big. Data. Comp. Vis. Vol. 4, No. 1 (2024) 22–30.

Paper Type: Original Article

Machine Learning-Based Authentication of Banknotes: A Comprehensive Analysis

Nadia Ghasem Abadi*

Department of Computer Engineering, Ayandegan Institute of Higher Education, Tonekabon, Iran; ghasemabadinadia@gmail.com.

Citation:

Received: 12 November 2023 *Revised: 14 Januray 2024* *Accepted: 23 Februray 2024*	Ghasem Abadi, N. (2024). Machine learning-based authentication of banknotes: a comprehensive analysis. *Big data and computing visions, 4*(1), 22-30.

Abstract

This research investigates the utilization of machine learning techniques for the identification and classification of counterfeit currency. The study utilizes a dataset consisting of authentic and counterfeit banknotes, employing various classification algorithms to construct a robust model for automated detection. Key features, including texture, color distribution, and security attributes, are extracted to train the model, enabling a thorough analysis of banknote authenticity. The proposed system exhibits promising accuracy in distinguishing genuine currency from counterfeits, thereby enhancing security measures in financial transactions and mitigating economic fraud.

Keywords: Fake currency, Counterfeit detection, Machine learning, Banknote authenticity, Economic fraud prevention.

1 | Introduction

The issue of Fake Currency Detection poses a significant and ongoing threat to individuals and businesses alike. With counterfeiters persistently devising novel methods and techniques, counterfeit banknotes have become increasingly indistinguishable from genuine currency, at least to the naked eye. In this article, we delve into the realm of fake currency detection with machine learning, a powerful approach that leverages the capabilities of artificial intelligence to combat this pervasive problem. Fake Currency Detection, in essence, involves the task of binary classification within the realm of machine learning. By harnessing a substantial dataset comprising both authentic and counterfeit banknotes, it becomes possible to train a model capable of accurately discerning the authenticity of new banknotes. This technology holds tremendous potential for bolstering the security measures associated with financial transactions and safeguarding against the detrimental effects of economic fraud [1], [2].

Through the utilization of machine learning algorithms, the detection of counterfeit currency transcends the limitations of human perception. While visual identification alone may prove challenging due to the high

Corresponding Author: ghasemabadinadia@gmail.com

https://doi.org/10.22105/bdcv.2024.197120

quality of counterfeit banknotes, machine learning models can extract intricate patterns, subtle variations, and distinguishing features from the data. By discerning unique characteristics such as texture, color distribution, security features, and other relevant attributes, these models can effectively classify banknotes as either genuine or counterfeit. The availability of a robust and accurate fake currency detection system has far-reaching implications for various stakeholders. For individuals, it ensures protection against financial losses resulting from unknowingly accepting counterfeit currency [3]. Businesses, on the other hand, can mitigate risks associated with counterfeit transactions and maintain the integrity of their financial operations. Moreover, the implementation of such systems contributes to the overall stability of the economy by curbing the proliferation of counterfeit banknotes and deterring economic fraud [4].

In this article, we aim to present a comprehensive overview of fake currency detection with machine learning, encompassing various methodologies, techniques, and advancements in the field. We will explore the significance of a diverse and representative dataset, the selection and optimization of machine learning algorithms, as well as the evaluation metrics employed to assess the performance of detection models. Additionally, we will discuss the practical implications and potential challenges associated with the deployment of these systems in real-world scenarios [5]. By shedding light on the capabilities and potential of machine learning-based fake currency detection, this article aims to contribute to the advancement of security measures in financial transactions, protect individuals and businesses from counterfeit currency, and foster a more resilient and fraud-resistant economic landscape [6].

2 | Literature Review

The literature review on fake currency detection with machine learning encompasses a comprehensive examination of existing research in the field [7]. Numerous studies have explored various methodologies, ranging from traditional image processing techniques to advanced machine learning algorithms.

2.1 | Traditional Approaches

Traditional approaches in fake currency detection primarily relied on image processing techniques. These techniques involved methods such as watermark detection, microprinting analysis, and UV feature extraction. While these approaches had some level of effectiveness, they often struggled to adapt to the evolving techniques used by counterfeiters [8]. As counterfeiters became more sophisticated, these traditional methods became less reliable in distinguishing between genuine and counterfeit banknotes.

2.2 | Machine Learning in Currency Authentication

The application of machine learning techniques in fake currency detection has gained considerable attention in recent years. Researchers have explored various machine learning algorithms to automate the process of authenticating banknotes and distinguishing them from counterfeits.

Support Vector Machines (SVMs), Random Forests, and Neural networks are among the commonly used machine learning algorithms in this domain. These algorithms are trained on a dataset of known authentic and counterfeit banknotes, where they learn to identify patterns and characteristics that differentiate between genuine and fake currency. SVM is a popular algorithm that excels in binary classification tasks. It constructs a hyperplane in a high-dimensional feature space to separate the data points into different classes [9]. In the context of fake currency detection, SVM can learn to separate authentic banknotes from counterfeits based on the extracted features. It aims to find the optimal decision boundary that maximizes the margin between the classes, resulting in improved classification accuracy.

Random Forests is an ensemble learning method that combines multiple decision trees to make predictions. Each decision tree is trained on a random subset of the data and features, and the final prediction is determined by aggregating the predictions of individual trees. In the context of fake currency detection, Random Forests can leverage the diversity of decision trees to capture various discriminative features and make accurate predictions.

Neural networks, particularly deep Neural networks, have gained significant attention in various fields, including fake currency detection. These networks consist of multiple layers of interconnected nodes (neurons) that learn to extract and model complex relationships in the data. Neural networks can be trained to process banknote images or extracted features and make predictions on their authenticity. By leveraging their ability to learn hierarchical representations and capture non-linear relationships, Neural networks can achieve high accuracy in counterfeit detection [9].

The training process of these machine learning models involves feeding them with labeled data, consisting of authentic and counterfeit banknotes, along with their corresponding features. The models learn to map the input features to the correct banknote class during the training phase. Once trained, these models can be used to classify new, unseen banknotes as genuine or counterfeit based on their learned patterns and features. It is worth noting that the choice of machine learning algorithm depends on several factors, including the complexity of the problem, the availability and quality of data, and the computational resources [10], [11]. Researchers often experiment with different algorithms and evaluate their performance using metrics such as accuracy, precision, recall, and F1-score to identify the most effective approach for counterfeit currency detection.

2.3 | Feature Extraction and Selection

Feature extraction plays a crucial role in enhancing the accuracy of counterfeit currency detection models. Researchers have focused on extracting intricate features from banknotes, such as security thread patterns, hologram characteristics, and microtext analysis. These features provide unique and distinguishing information that can be used by machine learning algorithms to classify banknotes accurately [12]. Additionally, feature selection methodologies are employed to identify the most relevant and informative attributes for classification, further improving the performance and efficiency of the detection models [13].

2.4 | Integration of Deep Learning

The integration of deep learning techniques has emerged as a notable trend in the field of fake currency detection. Deep learning models, particularly Convolutional Neural networks (CNNs) and Recurrent Neural networks (RNNs), have shown great promise in improving the accuracy and robustness of counterfeit detection systems. CNNs are widely recognized for their effectiveness in image analysis tasks. These models are designed to automatically learn hierarchical representations of image features. In the context of fake currency detection, CNNs can be trained to analyze visual characteristics of banknotes, such as patterns, textures, and security features.

By employing multiple layers of convolutional filters, CNNs can capture intricate details and discern subtle differences between genuine and counterfeit banknotes. This ability to learn complex image features makes CNNs well-suited for accurately identifying counterfeit patterns and distinguishing them from genuine banknotes. Recurrent Neural Networks (RNNs) are another class of deep learning models that excel in processing sequential data. In the context of fake currency detection, RNNs can be employed to analyze sequential patterns associated with security features or specific banknote attributes [14]. For example, RNNs can be trained to identify and analyze the sequence of hologram characteristics or the arrangement of microtext on a banknote. By capturing the temporal dependencies within the sequential data, RNNs can effectively detect counterfeit banknotes that exhibit anomalies or inconsistencies in their sequential patterns.

The integration of deep learning techniques into fake currency detection offers several advantages. Firstly, deep learning models can automatically learn relevant features from raw data, eliminating the need for manual feature engineering. This ability to learn discriminative features directly from the data enhances the adaptability and generalization capabilities of the detection models. Additionally, deep learning models can handle complex and non-linear relationships between the input data and the target labels, enabling them to capture intricate patterns that may not be easily discernible using traditional methods.

Nevertheless, there are challenges associated with the integration of deep learning techniques. Deep learning models often require large amounts of labeled training data to achieve optimal performance. Acquiring a diverse and representative dataset of counterfeit banknotes can be challenging due to their limited availability. Furthermore, deep learning models are computationally intensive and may require substantial computational resources for training and inference. Model optimization, regularization techniques, and hardware acceleration methods are commonly employed to mitigate these challenges and improve the efficiency of deep learning models in fake currency detection.

2.5 | Real-World Implementations

Some studies highlight practical implementations of fake currency detection systems in financial institutions and ATMs. Real-world scenarios provide insights into the feasibility, scalability, and effectiveness of deploying machine learning solutions for large-scale currency authentication.

3 | Fake Currency Detection

In the field of currency authentication, the detection of counterfeit banknotes plays a crucial role in maintaining the integrity of financial systems. To address this issue, a dataset specifically designed for fake currency detection is utilized in this study. Our dataset[1] comprises four key input characteristics that have been extracted from banknote images. These characteristics are as follows:

I. Variance of the image transformed into wavelets: this feature measures the variation or spread of the wavelet-transformed image. Wavelet transforms are commonly used in image analysis to capture both local and global variations present in an image.

II. Asymmetry of the image transformed into wavelets: the asymmetry feature quantifies the degree of symmetry or asymmetry in the wavelet-transformed image. It provides information about the distribution of intensity values across different regions of the banknote.

III. Kurtosis of the image transformed into wavelets: kurtosis is a statistical measure that describes the shape of the probability distribution of a dataset. In the context of fake currency detection, kurtosis of the wavelet-transformed image indicates the presence of outliers or extreme values in the image.

IV. Image entropy: entropy is a measure of the randomness or uncertainty in an image. In the context of banknote authentication, image entropy represents the level of complexity or information content present in the image. A higher entropy value suggests a more intricate pattern or texture within the banknote.

The analysis of these four input characteristics enables the application of machine learning algorithms and statistical techniques to create robust models that effectively identify counterfeit banknotes. The utilization of this dataset not only facilitates the development of research in the realm of fake currency detection but also strengthens the security of financial transactions.

A pair diagram is constructed to provide a comprehensive visualization of the relationships among the entities in the dataset. This diagram incorporates colored observations, with genuine banknotes depicted in blue and counterfeit banknotes represented in orange. By employing this visual representation, a clear overview of the associations between different entities can be obtained, aiding in the analysis of the dataset and facilitating further insights into the characteristics of genuine and counterfeit banknotes.

The pair plot provides valuable insights into the dataset, revealing several noteworthy observations:

[1] https://raw.githubusercontent.com/amankharwal/Website-data/master/data_banknote_authentication.txt

I. Variance and skewness exhibit distinct distributions for the two target characteristics, suggesting potential discriminative power in distinguishing genuine and counterfeit banknotes. In contrast, kurtosis and entropy exhibit similar distributions, indicating less discriminatory ability between the two classes.

II. The pair plot highlights both linear and nonlinear trends among the input features, indicating correlations and potential predictive relationships between certain characteristics.

III. Certain features demonstrate a clear separation between genuine and counterfeit banknotes, indicating their potential significance in accurately classifying banknotes.

To further assess the balance of the dataset concerning the target values, a distribution plot is generated (*Fig. 2*).

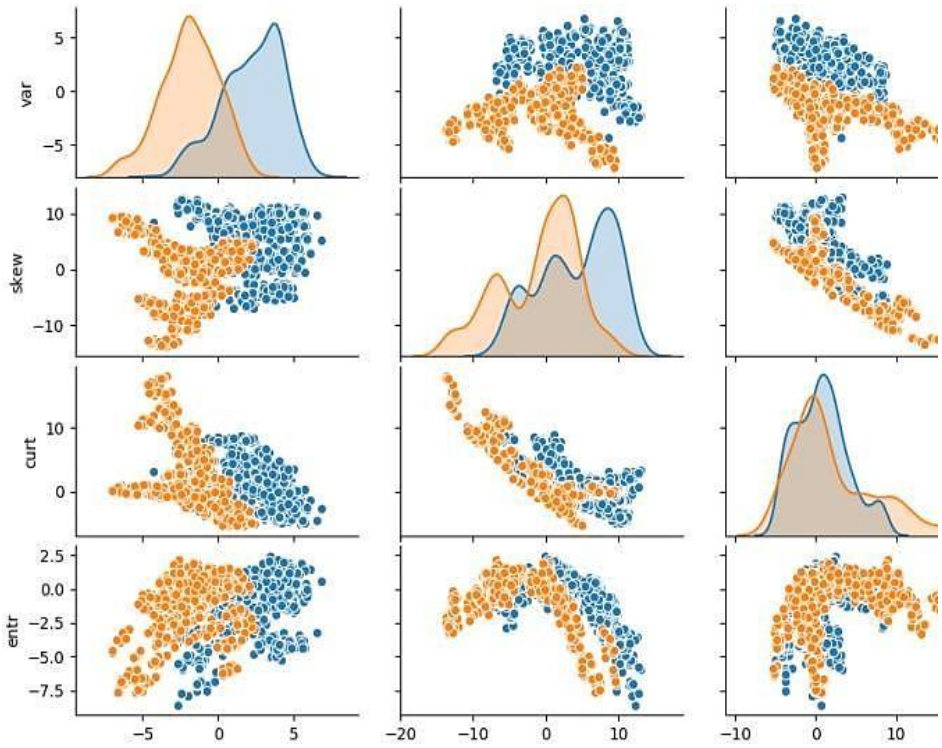

Fig. 1. Pair diagram illustrating the relationship between entities and color-coded observations of genuine (blue) and counterfeit (orange) banknotes.

The distribution plot *Fig. 2* illustrates the balance between the target values, denoted as 'auth'. The plot displays a count of the occurrences of each target value, with genuine and counterfeit banknotes represented by 0 and 1, respectively. The annotations indicate the specific counts for each category. The plot demonstrates the distribution of banknote samples across the two target values, providing an understanding of the dataset's class distribution.

This analysis contributes to the evaluation and understanding of the dataset, shedding light on the relationships between features, their discriminatory potential, and the distribution of target values.

The dataset exhibits a relatively balanced distribution; however, for the purpose of the binary classification task, achieving a perfect balance becomes imperative. Therefore, to ensure optimal model performance and mitigate any potential bias, preprocessing steps will be undertaken to address this requirement.

The initial preprocessing step involves carefully balancing the dataset to achieve an equal representation of both genuine and counterfeit banknotes. By attaining a balanced dataset, we can mitigate the risk of skewed classification results and improve generalization capabilities. This crucial preprocessing stage aims to create a dataset that accurately reflects the real-world distribution of genuine and counterfeit banknotes, enabling the subsequent classification model to make informed and unbiased predictions.

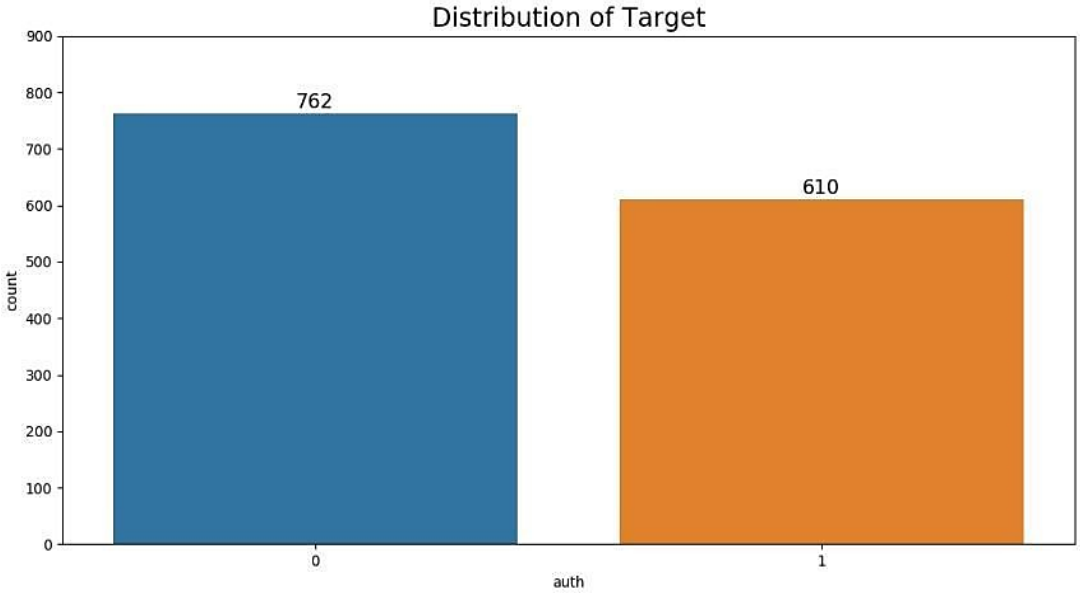

Fig. 2. Distribution of target values.

3.1 | Data Processing

To address the need for a balanced dataset, we employ random undersampling, which involves randomly dropping instances of the overrepresented target function. An alternative approach is oversampling, wherein synthetic data is created for the underrepresented target class. In this case, we will proceed with random undersampling by randomly removing 152 observations of genuine banknotes. To achieve this, we apply the following steps:

Step 1. Random undersampling. We randomly select observations from the dataset, ensuring a balanced representation of both genuine and counterfeit banknotes. By dropping a specific number of instances from the overrepresented target class, we create a balanced dataset. The resulting dataset, denoted as "Data" is obtained by randomly deleting 152 observations of genuine banknotes.

```
Nb_to_delete = target_count[0] - target_count[1]
Data = data.sample(frac=1, random_state=42).sort_values(by='auth')
Data = data[nb_to_delete:]
Print(data['auth'].value_counts())
```

Fig. 3. Random undersampling for class balancing.

The output of the above code snippet demonstrates the balanced dataset, where both genuine and counterfeit banknotes possess an equal count of 610 observations each.

Next, we divide the balanced dataset into training and test sets to facilitate model training and evaluation. The features are represented by 'x,' while the target variable, 'auth,' is denoted by 'y.' The dataset is split using the train_test_split function, with a test size of 0.3 and a random state of 42. By performing random undersampling and dividing the dataset into training and test sets, we ensure a balanced representation of both target classes and establish the necessary foundation for subsequent model development and evaluation.

Step 2. Data Standardization. To ensure data consistency and facilitate accurate model training, the StandardScaler method provided by Scikit-learn is employed for data standardization. The StandardScaler object is instantiated and fitted to the training set using the fit method. Subsequently, the transform method is applied to both the training and test sets, resulting in standardized feature values. This process is crucial for achieving optimal model performance by normalizing the feature scales and minimizing the influence of varying magnitudes.

4 | Logistic Regression for Fake Currency Detection

In this study, the task of fake currency detection is approached by leveraging the logistic regression algorithm. Logistic regression is a widely adopted technique for binary classification tasks due to its ability to estimate the probability of a specific class assignment. By modeling the dataset using logistic regression, the goal is to acquire a robust understanding of the underlying patterns and relationships that distinguish genuine banknotes from counterfeit ones. Logistic regression operates by fitting a logistic function to the training data, which maps the input features to the probability of belonging to a particular class. The algorithm optimizes its parameters through an iterative process that aims to minimize the discrepancy between predicted probabilities and the actual class labels [8]. By iteratively adjusting the model's coefficients, logistic regression learns to make informed decisions about class assignments.

The logistic regression model's strength lies in its interpretability, as the estimated coefficients provide insight into the influence of each feature on the classification outcome. This information can offer valuable insights into the characteristics that differentiate genuine banknotes from counterfeit ones.

5 | Model Training and Evaluation

After instantiating and fitting the logistic regression model, the subsequent step involves training the model using the provided dataset. The training data serves as the foundation for the model to learn the relationships between the input features and the corresponding target variable, which represents the authenticity of banknotes. During the model training phase, the logistic regression algorithm adjusts its parameters to minimize the discrepancy between predicted probabilities and the true class labels. This process involves an iterative optimization technique, such as maximum likelihood estimation or gradient descent, to find the optimal values for the model's coefficients.

By leveraging the fit method, the logistic regression model iteratively updates its coefficients based on the training data, continually improving its ability to accurately classify banknotes as genuine or counterfeit. The aim is to achieve a model that generalizes well to unseen data and exhibits robust performance in detecting fake currency. Through diligent model training, the logistic regression algorithm strives to capture the underlying patterns and relationships within the data, enabling it to make reliable predictions on new, unseen banknotes. The effectiveness of the trained model will be assessed and validated in subsequent evaluation and testing phases to ascertain its accuracy and reliability in fake currency detection.

To assess the model's performance, predictions are made on the test dataset using the predict method of the logistic regression model. The predicted values, denoted as Y_pred, are converted into a numpy array for further analysis. To evaluate the model's accuracy, a confusion matrix is constructed using the confusion_matrix function from the scikit-learn library. The confusion matrix provides an overview of the model's performance by comparing the predicted labels with the actual labels.

The confusion matrix is then displayed as a pandas DataFrame, named Conf_mat, with "Pred.Negative" and "Pred.Positive" as the column labels, and "Act.Negative" and "Act.Positive" as the row labels. Additionally, the individual values of true negatives (tn), false positives (fp), false negatives (fn), and true positives (tp) are extracted from the confusion matrix.

The accuracy of the model is calculated by dividing the sum of true negatives and true positives by the total number of observations in the test dataset. The resulting accuracy value is rounded to four decimal places.

The accuracy is then displayed alongside the confusion matrix, representing the model's ability to correctly classify banknotes.

Furthermore, it is worth noting that the model achieved an accuracy of 98.36%. Additionally, when the model predicted that a banknote was genuine, it was correct 100% of the time, indicating a high level of precision in identifying genuine banknotes.

Next, a simulation of a single banknote prediction is performed using the trained model. To simulate the prediction of a single banknote, the model requires the extraction of features from the new banknote. These features are then scaled using the previously fitted StandardScaler object, scalar. The scaled features are integrated into the pre-trained logistic regression model using the predict method. The prediction result is displayed, indicating the predicted class of the banknote. In this case, "Class0" signifies that the banknote is predicted to be non-genuine. Additionally, the probabilities of the banknote belonging to each target class are displayed. The probabilities [0/1] represent the likelihood of the banknote being classified as non-genuine (0) or genuine (1), respectively. In this example, the predicted probabilities are [0.61112576 0.38887424], indicating a higher likelihood of the banknote being classified as non-genuine.

Table 1. Confusion matrix and model performance metrics.

	Pred.Negative	**Pred.Positive**
Act.Negative	187	6
Act.Positive	0	173

Table 1 represents the confusion matrix, showing the number of instances classified as positive or negative by the model compared to the actual positive and negative instances in the test dataset. The accuracy of the model is calculated as 98.36%. The subsequent prediction indicates that the banknote is predicted to belong to Class0 (non-genuine), with a probability of 0.61112576 for Class0 and 0.38887424 for Class1 (genuine).

6 | Challenges and Limitations

During the implementation of the logistic regression model for fake currency detection, several challenges and limitations were encountered. One of the primary challenges involved feature selection and engineering. The selection of appropriate features that capture the distinguishing characteristics of genuine and counterfeit banknotes is crucial for model performance. Additionally, the availability and quality of the dataset can pose limitations, as insufficient or biased data may affect the model's ability to generalize to new instances.

Another challenge lies in addressing class imbalance, where one class (e.g., counterfeit banknotes) may be significantly underrepresented compared to the other class (e.g., genuine banknotes). Class imbalance can hinder model training and lead to biased predictions. Techniques such as oversampling or undersampling may be employed to mitigate this issue. Besides, it is important to acknowledge that the logistic regression model assumes a linear relationship between the features and the log-odds of the target variable. While this assumption may hold in some cases, complex non-linear relationships may exist in the data, which could limit the model's ability to capture intricate patterns accurately.

7 | Conclusion

In this study, the application of logistic regression for fake currency detection was explored. The logistic regression model demonstrated promising performance, achieving an accuracy of 98.36% on the test dataset. The model's ability to accurately classify banknotes as genuine or counterfeit holds significant potential in mitigating financial fraud and ensuring the integrity of monetary transactions. However, it is crucial to recognize the limitations and challenges associated with logistic regression as well. Addressing feature selection, class imbalance, and potential non-linear relationships are essential considerations for further improving the model's performance.

Overall, the findings of this research highlight the effectiveness of logistic regression in fake currency detection and provide a foundation for future investigations in developing more robust and accurate models for financial security applications.

References

[1] Abdulalem, A., Shukor, A. R., Siti Hajar, O., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., … & Abdu, S. (2022). Financial fraud detection based on machine learning: a systematic literature review. *Applied sciences*, *12*(19), 9637.

[2] Das Guptta, S., Shahriar, K. T., Alqahtani, H., Alsalman, D., & Sarker, I. H. (2024). Modeling hybrid feature-based phishing websites detection using machine learning techniques. *Annals of data science*, *11*(1), 217–242.

[3] Tang, L., & Mahmoud, Q. H. (2021). A survey of machine learning-based solutions for phishing website detection. *Machine learning and knowledge extraction*, *3*(3), 672–694.

[4] Mitra, A., Mohanty, S. P., Corcoran, P., & Kougianos, E. (2021). A machine learning based approach for deepfake detection in social media through key video frame extraction. *SN computer science*, *2*(2), 98.

[5] Diaz-Escobar, J., Ordonez-Guillen, N. E., Villarreal-Reyes, S., Galaviz-Mosqueda, A., Kober, V., Rivera-Rodriguez, R., & Rizk, J. E. L. (2021). Deep-learning based detection of COVID-19 using lung ultrasound imagery. *Plos one*, *16*(8), e0255886.

[6] Shoaib, M., Hussain, T., Shah, B., Ullah, I., Shah, S. M., Ali, F., & Park, S. H. (2022). Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease. *Frontiers in plant science*, *13*, 1031748.

[7] Abedi, V., Misra, D., Chaudhary, D., Avula, V., Schirmer, C. M., Li, J., & Zand, R. (2024). Machine learning-based prediction of stroke in emergency departments. *Therapeutic advances in neurological disorders*, *17*, 17562864241239108. https://doi.org/10.1177/17562864241239108

[8] Rahnamay Roodposhty, F., & Imeni, M. (2023). The historiography of mathematics for finance and accounting. *Journal of management accounting and auditing knowledge*, *12*(45), 121–134.

[9] Boulahia, S. Y., Amamra, A., Madi, M. R., & Daikh, S. (2021). Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition. *Machine vision and applications*, *32*(6), 121.

[10] Callegari, C., Giordano, S., & Pagano, M. (2024). A real time deep learning based approach for detecting network attacks. *Big data research*, 100446.

[11] Kuang, M., Safa, R., Edalatpanah, S. A., & Keyser, R. S. (2023). A hybrid deep learning approach for sentiment analysis in product reviews. *Facta universitatis, series: mechanical engineering*, *21*(3), 479–500.

[12] Huang, Z., Zheng, H., Li, C., & Che, C. (2024). Application of machine learning-based k-means clustering for financial fraud detection. *Academic journal of science and technology*, *10*(1), 33–39.

[13] Javadi, S., Safa, R., Azizi, M., & Mirroshandel, S. A. (2020). A recommendation system for finding experts in online scientific communities. *Journal of ai and data mining*, *8*(4), 573–584.

[14] Taghvaei, F., & Safa, R. (2021). Efficient energy consumption in smart buildings using personalized NILM-based recommender system. *Big data and computing visions*, *1*(3), 161–169.